**Documentation of Neospectra datasets**

February 14th, 2025

**Zimbabwe – Double Burden dataset**

Link to cleaned data with metadata: <https://github.com/FBaudron/Neospectra-analysis/blob/c5f41bc6161486edd710c636e80b6c484c7e918a/Data_Neospectra_Zimbabwe_Double_Burden_clean_with_metadata.xlsx>

The dataset contains selected soil properties from 236 soil samples (145 0-5 cm samples and 91 5-20 cm sample) subjected to wet chemistry, MIRS analysis, ICP-OES analysis, and portable NIRS (using a Neospectra device). Wet chemistry, MIRS analysis, and ICP-OES analysis were conducted by Rothamsted Research. All soils were scanned five times at the International Maize and Wheat Improvement Centre (CIMMYT) in Harare with a portable NIRS (Neospectra).

These soil samples were collected in 2020 from homefields and outfields of 30 farms, as well as well as from nearby pasture/savannah (“Virgin (no cultivation)” in the database). These farms were selected in Ward 4 and Ward 27 of Murehwa District through a stratified sampling strategy, using a farm typology described in Hassall et al. (2023). Each sample is a composite sample of 10-15 random points from each plot, collected with a soil auger.

The list of variables included in the dataset, their descriptions and their units are given below (metadata).

|  |  |  |
| --- | --- | --- |
| Variable | Description | Unit |
| RRES | Soil ID | None |
| Soil.depth | Depth of the soil layer | 0-5 cm; 5-20 cm |
| Land.Use | Land use where soil was sampled | Virgin (no cultivation); Homefield (field close to homestead): Outfield (field distant from homestead) |
| pH | pH in 1:2.5 soil:water suspension (ISO 10390: 2005) | None |
| Total.soil.N | Total nitrogen content determined by combustion (LECO, Michigan, USA) | % |
| Total.soil.C | Total carbon content determined by combustion (LECO, Michigan, USA) | % |
| OlsenP | P Olsen | ppm |
| Ca.ex | Exchangeable Ca using ICP-OES (Optima 7300 DV, Perkin Elmer, CT, USA) following cobalt hexamine extraction (ISO 23470: 2018) | cmolc/kg |
| K.ex | Exchangeable K using ICP-OES (Optima 7300 DV, Perkin Elmer, CT, USA) following cobalt hexamine extraction (ISO 23470: 2018) | cmolc/kg |
| Mg.ex | Exchangeable Mg using ICP-OES (Optima 7300 DV, Perkin Elmer, CT, USA) following cobalt hexamine extraction (ISO 23470: 2018) | cmolc/kg |
| Na.ex | Exchangeable Na using ICP-OES (Optima 7300 DV, Perkin Elmer, CT, USA) following cobalt hexamine extraction (ISO 23470: 2018) | cmolc/kg |
| eCEC.clean | Effective cation exchange capacity using ICP-OES (Optima 7300 DV, Perkin Elmer, CT, USA) following cobalt hexamine extraction (ISO 23470: 2018) | cmolc/kg |
| Base.sat | Base saturation (calculated) | % |
| clay.% | Clay content using MIR spectra with an internal calibration (Thomas et al., 2021) | % |
| silt.% | Silt content using MIR spectra with an internal calibration (Thomas et al., 2021) | % |
| sand.% | Sand content using MIR spectra with an internal calibration (Thomas et al., 2021) | % |
| Al | Al concentration using ICP-OES (Optima 7300 DV, Perkin Elmer, CT, USA) following aqua regia digestion (McGrath and Cunliffe, 1985) | ppm |
| B | B concentration using ICP-OES (Optima 7300 DV, Perkin Elmer, CT, USA) following aqua regia digestion (McGrath and Cunliffe, 1985) | ppm |
| Ca | Ca concentration using ICP-OES (Optima 7300 DV, Perkin Elmer, CT, USA) following aqua regia digestion (McGrath and Cunliffe, 1985) | ppm |
| Cd | Cd concentration using ICP-OES (Optima 7300 DV, Perkin Elmer, CT, USA) following aqua regia digestion (McGrath and Cunliffe, 1985) | ppm |
| Co | Co concentration using ICP-OES (Optima 7300 DV, Perkin Elmer, CT, USA) following aqua regia digestion (McGrath and Cunliffe, 1985) | ppm |
| Cr | Cr concentration using ICP-OES (Optima 7300 DV, Perkin Elmer, CT, USA) following aqua regia digestion (McGrath and Cunliffe, 1985) | ppm |
| Cu | Cu concentration using ICP-OES (Optima 7300 DV, Perkin Elmer, CT, USA) following aqua regia digestion (McGrath and Cunliffe, 1985) | ppm |
| Fe | Fe concentration using ICP-OES (Optima 7300 DV, Perkin Elmer, CT, USA) following aqua regia digestion (McGrath and Cunliffe, 1985) | ppm |
| K | K concentration using ICP-OES (Optima 7300 DV, Perkin Elmer, CT, USA) following aqua regia digestion (McGrath and Cunliffe, 1985) | ppm |
| Mg | Mg concentration using ICP-OES (Optima 7300 DV, Perkin Elmer, CT, USA) following aqua regia digestion (McGrath and Cunliffe, 1985) | ppm |
| Mn | Mn concentration using ICP-OES (Optima 7300 DV, Perkin Elmer, CT, USA) following aqua regia digestion (McGrath and Cunliffe, 1985) | ppm |
| Mo | Mo concentration using ICP-OES (Optima 7300 DV, Perkin Elmer, CT, USA) following aqua regia digestion (McGrath and Cunliffe, 1985) | ppm |
| Na | Na concentration using ICP-OES (Optima 7300 DV, Perkin Elmer, CT, USA) following aqua regia digestion (McGrath and Cunliffe, 1985) | ppm |
| Ni | Ni concentration using ICP-OES (Optima 7300 DV, Perkin Elmer, CT, USA) following aqua regia digestion (McGrath and Cunliffe, 1985) | ppm |
| P | P concentration using ICP-OES (Optima 7300 DV, Perkin Elmer, CT, USA) following aqua regia digestion (McGrath and Cunliffe, 1985) | ppm |
| Pb | Pb concentration using ICP-OES (Optima 7300 DV, Perkin Elmer, CT, USA) following aqua regia digestion (McGrath and Cunliffe, 1985) | ppm |
| S | S concentration using ICP-OES (Optima 7300 DV, Perkin Elmer, CT, USA) following aqua regia digestion (McGrath and Cunliffe, 1985) | ppm |
| Se | Se concentration using ICP-OES (Optima 7300 DV, Perkin Elmer, CT, USA) following aqua regia digestion (McGrath and Cunliffe, 1985) | ppm |
| Ti | Ti concentration using ICP-OES (Optima 7300 DV, Perkin Elmer, CT, USA) following aqua regia digestion (McGrath and Cunliffe, 1985) | ppm |
| Zn | Zn concentration using ICP-OES (Optima 7300 DV, Perkin Elmer, CT, USA) following aqua regia digestion (McGrath and Cunliffe, 1985) | ppm |
| Sample.Name | Neospectra sample ID | None |
| Device.Id | Neospectra device ID | None |
| Created.At.(UTC) | Date and time Neospectra scanning | Date time |
| Created.By | Email of Neospectra user | None |
| 2549.9999824259398 to 1350.1554633964199 | Reflectance at difference wavelength (in nm) | % |

**Ethiopia – GAIA dataset**

Link to cleaned data with metadata: <https://github.com/FBaudron/Neospectra-analysis/blob/c5f41bc6161486edd710c636e80b6c484c7e918a/Data_Neospectra_Ethiopia_GAIA_clean_with_metadata.xlsx>

The dataset contains selected soil properties predicted (and calculated) from 120 soil samples (44 0-20 cm samples from Jimma, 40 0-20 cm samples from Debre Markos, and 36 20-50 cm samples from Debre Markos) subjected to MIRS analysis and portable NIRS (using a Neospectra device). MIRS was conducted by the International Centre for Research in Agroforestry (ICRAF) in Kenya, and predictions were made from a machine-learning algorithm, using results from wet chemistry analysis (conducted on 10% of the soil samples at the CROPNUTS Laboratory Services in Kenya). All soils were scanned five times at the International Maize and Wheat Improvement Centre (CIMMYT) in Addis Ababa with a portable NIRS (Neospectra).

These soil samples were collected in 2021 from farmers’ fields selected through a spatial sampling strategy, using two strata of cropland: (1) areas with pH below 5.5 and (2) areas with pH between 5.5 and 6.5, as identified from ISDA spatial data. In each farmers’ field, five topsoil and subsoil soil samples were collected with an auger and bulked into one 0-20 cm composite sample, and one 20-50 cm depth composite sample.

The list of variables included in the dataset, their descriptions and their units are given below (metadata).

|  |  |  |
| --- | --- | --- |
| Variable | Description | Unit |
| Neospectra | Neospectra sample ID | None |
| ICRAF | Wet chemistry Sample codes for ICRAF | None |
| Spectral | Spectral sample code for ICRAF | None |
| pH | pH in 1:2 soil:water suspension (predicted) | None |
| SOC | Soil organic carbon (predicted) | % |
| Clay | Clay content (predicted) | % |
| TN | Total nitrogen content (predicted) | % |
| m3.Al | Al concentration following Mehlich extraction (predicted) | ppm |
| m3.B | B concentration following Mehlich extraction (predicted) | ppm |
| m3.Ca | Ca concentration following Mehlich extraction (predicted) | ppm |
| m3.Fe | Fe concentration following Mehlich extraction (predicted) | ppm |
| m3.Mg | Mg concentration following Mehlich extraction (predicted) | ppm |
| m3.Mn | Mn concentration following Mehlich extraction (predicted) | ppm |
| m3.Na | Na concentration following Mehlich extraction (predicted) | ppm |
| m3.S | S concentration following Mehlich extraction (predicted) | ppm |
| ExAc | Exchangeable acidity (predicted) | cmolc/kg |
| m3.K | K concentration following Mehlich extraction (predicted) | ppm |
| PSI | Phosphorus Sorption Index (predicted) | None |
| Sand | Sand content (predicted) | % |
| CEC | Cation Exchange Capacity (predicted) | cmolc/kg |
| Estimated.Acid.Saturation.... | Acid saturation (calculated) | % |
| Estimated\_Silt | Silt content (calculated) | % |
| Soil\_Textural\_Class | USDA soil-textural class | None |
| eCEC | Effective Cation Exchange Capacity (calculated) | cmolc/kg |
| Alsat | Aluminium saturation (calculated) | % |
| 3921.5686540000002 to 7392.9412570000004 | Reflectance at difference wavelength (in nm) | % |